CLEAN VERSION

METHOD FOR FABRICATING AN OLFACTORY RECEPTOR-BASED BIOSENSOR

This application is a continuation-in-part of application 09/057,181 filed April 8, 1998, the entire disclosure of which is incorporated herein by reference.

FIELD OF THE INVENTION

The present invention related generally to biosensors and, more specifically, to biosensors which have biomolecules attached to a thin film transducer.

BACKGROUND OF THE INVENTION

Chemoreception is an ancient sense system that enables organisms to detect chemicals in its environment. In humans, odor receptor cells are located in the nose. The biochemical receptors for the odorants are transmembrane proteins found in the membrane of receptor cells cilia. Olfactory receptor proteins (ORP) generally have seven non-intersecting helices. It is believed that conserved residues determine the orientation of each helix relative to the other helices.

The detection of environmental chemicals is mediated by peripheral olfactory organs of varied complexity in almost all metazoans. Typically, specialized sensory neurons initiate perception by detecting ambient molecules, commonly called odors,

that interact with protein receptors in their membranes. ORP on the cilia detect the odorants entering the nose. The ORP are coded by approximately 1000 genes, and it is the largest gene family in the genome of any species. ORP are members of the proteins having seven transmembrane domains, i.e. G-protein couple receptor (GPCR) superfamily. They have a diverse amino acid sequence and are able to recognize a wide variety of structurally diverse odorants. The amino acid sequences of ORP are especially variable in the several transmembrane domains. This is a possible mechanism for the recognition of a variety of structurally diverse ligands.

The major path of olfactory transduction is shown in Fig. 1. Binding of the odor molecules to the receptors may activate a G-protein coupled enzymatic cascade to generate second messengers. These messengers can open the ion channels on the membrane of olfactory cells. The opening channels may depolarize the membrane and lead to action potentials and signaling.

There is currently a need for sensors which function like an ORP being capable of detecting ligands, i.e. certain gas molecules, to be developed. The goal, then, is to develop useful sensors for detecting the presence of certain gas molecules according to the assignment of the certain gas molecules binding to certain sites of an ORP. It has been difficult in the past, however, to rapidly determine the secondary and tertiary molecular structures of ORPs having olfactory receptor binding domains specific to selected ligands of interest. This is due in part to the complexity of ORP molecules. As understood by those skilled in the art, in an empirical analysis, a determination of putative binding domains is an extremely labor-intensive endeavor. It begins with identification and molecular cloning of

genes that code for the receptor protein of interest. These genes are then expressed and the target protein is isolated and purified. Physical studies such as X-ray diffraction, neutron diffraction and electron microscopy are conducted to determine 2-D maps and 3-D structure; site directed mutagenesis is conducted to determine the position of residues for ligand binding. It would be desirable to provide a method which eliminates as many of these steps as possible.

SUMMARY OF THE INVENTION

In one aspect, the present invention provides a method for rapidly determining ORP candidates for use as receptors for preselected odorant molecules.

In another aspect, the present invention provides a method for fabricating a biosensor which includes a layer of peptides that selectively binds a preselected odorant molecule.

Accordingly, the present invention provides a method for making a biosensor capable of detecting a gas molecule, wherein the gas molecule is a ligand capable of binding to an olfactory receptor protein. The method includes the steps of determining the amino acid sequence of a preselected olfactory receptor protein the secondary and tertiary structures of which are not known. Typically this step will be carried out by choosing an ORP from a database of ORPs which have been sequenced. In the next step the ammo acid sequence of the ORP selected in the first step is compared to the sequence of G-coupled protein receptors having known secondary and tertiary structures. This step will typically be carried out by accessing a database of G-protein receptors having known primary, secondary and

tertiary structures. Next, based on primary sequence homology, one or more G-coupled protein receptors are chosen as a candidate on which to predict the secondary and tertiary structure of the unknown ORP. In the next step, the secondary and tertiary structures of the unknown ORP are approximated based on the known structures of the G-protein receptor selected through sequence homology comparison in the prior steps. The approximated secondary and tertiary structures of the unknown ORP are then analyzed using conventional modeling techniques to identify likely binding domains for the ligand of interest. A peptide is then synthesized having the primary sequence of the most likely binding domain for the ligand. These peptides are attached to a transducer. The resultant biosensor is then tested by exposing it to the target ligand and determining binding efficiencies.

By identifying and testing a number of peptides in this manner, high affinity biosensors can be rapidly fabricated.

BRIEF DESCRIPTION OF HE DRAWINGS

Fig. 1 is a diagram illustrating the major pathway of olfactory transduction.

Fig. 2 is a flow chart illustrating the modeling steps of the present invention.

Fig. 3 is an amino acid sequence for OLFD_CANFA (P30955).

Fig. 4 is a three dimensional structure showing the simulation results of the olfactory receptor protein, OLFD_CANFA (P30955), docking with trimethylamine which is shown as spherical molecular models.

Fig.5 is a perspective view of a transducer made in accordance with a preferred embodiment of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The detailed embodiment of the present invention is disclosed herein. It should be understood, however, that the disclosed embodiment is merely exemplary of the invention, which may be embodied in various forms. Therefore, the details disclosed herein are not to be interpreted as limited, but merely as the basis for the claims and as a basis for teaching one skilled in the art how to make and/or use the invention.

Fig. 2 is a flow chart illustrating the modeling steps of the preferred embodiment. Referring now to Fig. 2 of the drawings, an olfactory receptor protein which has been sequenced is selected in step 210. Of course, it may be desirable in some cases to actually clone, express, isolate and sequence a new ORP; however, in most instances an ORP will be chosen from a sequence database having the primary amino acid sequence of various ORPs. One preferred database for use in the present invention is available on the ExPASy server of the Swiss Institute of Bioinformatics. Other similar databases or print sources may be equally suitable.

Once the ExPASy server has been accessed, the "SWISS PROT and TrEMBL" database is opened. The ExPASy server is open to the public and may be accessed via the Internet. Next, using the keyword search features of this file, the key words "olfactory receptor" may be used to create a subset of sequences of olfactory receptor proteins. An ORP is then selected, the sequence of which is to be used in the practice of the invention. The known sequence is displayed along with additional information on the ORP such as EMBL cross references, length and

molecular weight. The amino acid sequence information is generally subdivided into potential extracellular, transmembrane and cytoplasmic domains, which are predicted and provided only for reference. For example, an ORP, OLFD_CANFA (P30955), is selected from the "SWISS PROT and TrEMBL" database. The amino acid sequence is shown on Fig. 3, and the predicted secondary-structure features of OLFD_CANFA (P30955) are listed in Table 1.

Table 1

Key	Position	Length	Description
Domain	1-25	25	Extracellular (potential)
Transmem	26-49	24	1 (potential)
Domain	50-57	8	Cytoplasmic (potential)
Transmem	58-79	22	2 (potential)
Domain	80-100	21	Extracellular (potential)
Transmem	101-120	20	3 (potential)
Domain	121-139	19	Cytoplasmic (potential)
Transmem	140-158	19	4 (potential)
Domain	159-195	37	Extracellular (potential)
Transmem	196-218	23	5 (potential)
Domain	219-235	17	Cytoplasmic (potential)
Transmem	236-259	24	6 (potential)
Domain	260-271	12	Extracellular (potential)
Transmem	272-291	20	7 (potential)
Domain	292-313	22	Cytoplasmic (potential)

In step 220 of Fig. 2, the predicted secondary structure, such as α-helix, β-sheet, and transmembrane regions, of the ORP under investigation is determined by

using, for example, the "PredictProtein" server of the "BIOcomputing 3D Modeling Unit Service" (B Rost: PHD: predicting one-dimensional protein structure by profile based neural networks. Methods in Enzymolgy, 266, 525-539, 1996). The "PredictProtein" server can be accessed through worldwide web sites. The service of "PredictProtein" includes sequence analysis and structure prediction. One can submit any protein sequence, and then "PredictProtein" retrieves similar sequences in the database and predicts aspects of protein structure. The "PredictProtein" server uses several programs and database, such as those listed in Table 2, to predict protein's structure.

Table 2

Program's Type	Program	Function
Alignment and database searching methods	MaxHom	MaxHom is a dynamic multiple sequence alignment program which finds similar sequence in a database.
Sequence motif searching methods	ProSite	ProSite is a database of functional motifs.
	ProDom	ProDom is a database of putative protein domains; searched with BLAST for domains corresponding to sequence being investigated.
Prediction of protein structure	PHDsec	PHDsec predicts secondary structure from multiple sequence alignments.
	PMDacc	PMDacc predicts per residue solvent accessibility from multiple sequence alignments.
	PHDhtm	PHDhtm predicts the location and topology of transmembrane helices from multiple sequence alignments.
	GLOBE	GLOBE predicts the globularity of a protein.

	TOPITS	TOPITS is a prediction-based threading program, that finds remote structural homologues in the DSSP database.
	COILS	COILS finds coiled-coil regions in your protein.
	EvalSec	EvalSec evaluates secondary structure prediction accuracy.

In essence, these servers allow the sequence of the ORP to be submitted for comparison to the sequences of proteins in the PredictProtein database. PredictProtein retrieves similar sequences and predicts secondary protein structure based on data for similar sequences. PredictProtein performs and displays the results of a "PROSITE" motif search, "ProDom" domain search, MAXHOM alignment header analysis, and provides information regarding accuracy of the forgoing analyses. This prediction of secondary structure is performed by PredictProtein using a system of neural networks.

The MAXHOM program produces a multiple sequence alignment file which serves as the input for the neural network system. The output of the MAXHOM analysis includes identification of aligned proteins, percentage of pairwise sequence identity, percentage of weighted similarity, number of residues aligned, number of insertions and deletion (indels), number of residues in all indels, length of aligned sequences and a short description of the aligned proteins. The preferred neural network for prediction of secondary structure is described in "Prediction of Protein Structure at Better than 70% accuracy" J. Mol. Bioi., 1993, 232, 584-599, and the entire disclosure of which is incorporated by reference.

Prediction of solvent accessibility is also determined (PHDacc) in accordance

with "The Analysis and Prediction of Solvent Accessibility in Protein Families" Proteins, 1994, 20, 216-226, and the entire disclosure of which is incorporated by reference. The latter prediction provides values for the relative solvent accessibility. Prediction of helical transmembrane segments of the ORP is performed by the PHDhtm program. In this manner, the secondary structure (helix, sheet, loop) and location relative to the membrane (inside, outside, transmembrane) for the ORP under investigation is predicted with relative accuracy. Most preferably, the predicted topology for the transmembrane proteins is determined using PHDtopology and fold recognition is determined by predicted-based threading using PHDthreader. Again, the secondary structure predictive determinations are verified for accuracy using EvalSec. All of the computer programs used in the present invention can be accessed by the public, and their disclosures are incorporated herein by reference.

For example, primary amino acid sequence of OLFD_CANFA (P30955) is input into the "PredictProtein" server. Since most of odorant molecules bind to transmembrane helices of an ORP, the predicted seven transmembrane helices of the OLFD_CANFA (P30955) are listed in Table 3.

Table 3

Number of helix	Sequence	Position of amino acids
1	FYALFLAMYVTTILGNLLIIVLIQ	27-50
2	LHTPMYLFLSNLSFSDLCFSSV	55-76
3	LTQMYFFLFFGDLESFLLVAMAYD	98-121
4	CFSLLVLSWVLTMFHAVLHTLLM	141-163

5	VIFIMGGLILVIPFLLIITSYARIV	197-221
6	SHLSVVSLFYGTVIGLYL	242-259
7	MAMMYTVVTPMLNPFIYS	273-290

In Fig. 2, after determining the predicted seven transmembrane helices, a template protein used to predict the approximated tertiary structure of the transmembrane helices are selected in step 230. This is preferably achieved in the preferred embodiment using the Swiss-Model interface program and, preferably, BLAST (Basic local alignment search tool as described in J. Mol. Biol. 215:403-410, the entire disclosure of which is incorporated herein by reference). To begin, the complete sequence of the ORP under investigation is input through Swiss-Model interface, and then the BLAST program determines the most appropriate modeling template to be used in the tertiary structure investigation. The modeling template will be that protein (of known primary, secondary and tertiary structures) having the highest primary sequence homology and similar secondary structure with the ORP to be investigated.

For example, the primary amino acid sequence of the ORP, OLFD_CANFA (P30955), is input through the Swiss-Model interface. The primary sequence of the OLFD_CANFA (P30955) is compared to the sequences of proteins in the 7TM (seven transmembrane) subset of the SWISS-PROT database by the BLAST program, since OLFD_CANFA (P30955) also has seven transmembrane helices. Then, a number of BALST-assisted templates, as listed in Table 4, are obtained. In Table4, Neuropeptide Y1 receptor (P25929) has the largest P(N). That is,

Neuropeptide Y1 receptor (P25929) has the highest primary sequence homology with the OLFD_CANFA (P30955). Hence, Neuropeptide Y1 receptor (P25929) is selected to be the modeling template of OLFD_CANFA (P30955).

Table 4

SWISS-PROT Code	Seven helices modeling template	Smallest Poisson Probability	
		P(N)	N
P25929	Neuropeptide Y1 receptor (*Homo sapiens*)	42	6.1×10^{-2}
P07550	Beta-2 adrenergic receptor (*Homo sapiens*)	37	2.8×10^{-1}
P21452	Substance-K receptor (Neurokinin A receptor)	39	7.0×10^{-4}
P02699	Rhodopsin (*Bos Taurus*)	41	5.1×10^{-8}
P02945	Bacteriorhodopsin (*Halobacterium halobium*)	*NA	*NA

*NA: not available.

After the modeling template has been selected, the sequences of the helical regions are displayed and the sequences of the helices of the ORP under investigation (as determined in the secondary structure analysis step of the present invention) are input through Swiss-Model interface program in step 240. That is, the

helical regions of the template are aligned with the helical regions of the ORP under investigation. The comparison yields a prediction of the tertiary structure (3D in space) of the ORP being investigated on an atom-by-atom basis. The tertiary structure of the ORP being under investigated is preferably output as a file containing three coordinates of each atom in the ORP. For example, a lengthy list of three coordinates of each atom in the OLFD_CANFA (P30955) was obtained.

The preferred protocol taken into consideration for the step 240 includes energy minimization and the like as described in: ProMod and Swiss-Model: Internet-based Tools for Automated Comparative Protein Modeling, Biochem. Soc. Trans. V. 24 274 1996; Large-Scale Comparative Protein Modeling, Proteome Research: New Frontiers in Functional Genomics 177 1997; Swiss-Model and the Swiss-PDBviewer; an Environment for Comparative Protein Modeling, Electrophoresis, V. 18 2714 1997; Automated Modeling of the Transmembrane Region of G-Protein Coupled Receptor by Swiss-Model, Receptors; and Channels v. 4 161 1996; Protein Modeling by email, Bio/Technology V. 13 658 1995, the disclosures of which are incorporated by reference.

The preferred modeling software programs which can be used in the present invention have a high degree of sophistication. For example, ProMod, which is under SWISS-MODEL Repository of the ExPASy Molecular Biology Server, applies a Protein Modelling tool which requires similarities with experimentally determined protein structures. It is a knowledge-based approach to predictive structure determination. It requires at least one known 3D structure of a related protein and good quality sequence alignments; the degree of sequence identity affects the accuracy of the predictive structure. In ProMod, there is a superposition of related

3D structures. A multiple alignment with the sequence under investigation is made. A framework for the new sequence is made and any missing loops are rebuilt. The backbone of the structure is completed and corrected if required. Side chains are corrected and rebuilt. The resultant structure is verified and packing is checked. The structure is then refined by energy minimization and molecular dynamics considerations.

The tertiary structures of the helices of the ORP under investigation are thus determined in step 240 and may be viewed stereoscopically using a program such as Insight II, a commercial program provided by Molecule Simulations Inc. and now is provided by Accelrys Inc., Swiss PDB-viewer or the like. Next, in step 250, a ligand, i.e. a gas is selected. A number of assays may be used to determine high general binding affinities of various ligands for the ORP under investigation. The molecular structure of the ligand and the ORP under investigation is then input to the Insight II program, i.e. the tertiary or 3D structures of ORP helices and the ligand are input. Next in step 260, the most probably geometrical binding domains of the ORP under investigation and the ligand are determined, preferably using the Global Range Molecular Modeling program (GRAMM) by geometric recognition algorithms. As understood by those skilled in the art, GRAMM is a program for protein docking, and it treats the ORP and the ligand as rigid bodies. Since GRAMM utilizes geometric recognition algorithms to determine the most probably geometrical binding domains of a protein for a ligand, no specific information about the binding sites is required. It performs a six-dimensional search through the relative translations and rotations of molecules. It takes an empirical approach to smoothing the intermolecular energy function by changing the range of atom-atom potentials. It allows the user to locate

the area of the global minimum of intermolecular energy for structures of different accuracy.

Then in step 270, the structures of the ORP and the ligand are allowed to relax. That is, the structures of the ORP and the ligand are flexible. Hence, the bond stretching, valence angle bending, torsion, van der Waals force, and electrodtatic force of both the ORP and the ligand are taken into consideration. Affinity Docking program, an embedded program of Insight II, is preferably used to calculate the energy distribution and reaction forces between the ligand and the geometrically biniding domains, as predicted by GRAMM, of the ORP by molecular mechanics calculations using the energy minimization algorithm. The most probably overall binding domains are thus determined, and the user can read out the sequence of the binding domains by move the mouse to each amino acid of the binding domains.

For example, the most probably binding domains, as shown in Fig. 4, of the OLFD_CANFA (P30955) for trimethylamine is predicted. The trimethylamine molecules are shown as spherical molecular model, and the OLFD_CANFA (P30955) is shown as cartoon structure. The eight most probably binding domains of the OLFD_CANFA (P30955) for the trimethylamine are located in transmembrane 1, transmembrane 3, and transmembrane 5.

Peptides are then synthesized corresponding to these most probably binding domains using conventional synthesis technologies. The peptides are then applied to the surface of a transducer, preferably one fabricated using thin film (semiconductor) technique as will be known to those skilled in the art. Briefly, with reference to Fig. 5, transducers 510 coated with peptide layer 520 are on biosensor 500. Transducer 510 is preferably a piezoelectric quartz crystal-based device. A

new change will occur if a ligand binds to the peptide layer resulting any measurable frequency change in the quartz crystal frequency, allowing detection of ligand binding. The success and efficiency of the transducer can be determined, including by comparing the sensor's response to the ligand and other molecules.

For example, peptides synthesized according to the most probably binding domains of OLFD_CANFA (P30955) for trimethylamine are peptides B1, B2, and B3. The amino acid sequences of the peptides B1, B2, and B3 are DPDQRDC, GDLESFC, and CFFLFFGD. These peptides all have or are added a cystein (symbol C) residue at one terminal. The transducers of a biosensor have gold electrodes, the –SH functional group of the cystein can react with gold electrodes directly in an organic solution to form chemical bond between them. Hence, a simple way to attach these peptides is dipping the surface of gold electrode on piezoelectric quartz with the peptide solution under room temperature for a period of time. Therefore, these peptides can be attached on the surfaces of the transducers. Then, after attaching these peptides on the transducers of a biosensor, the biosensor is used to detect various gases such as trimethylamine, dimethylamine, ammonia, acetone, formic acid, ethanol, and formaldehyde. The experimental results of the peptides B1, B2, B3, and Pb2 are listed in Table 6, wherein the peptide Pb2 is not designed according to the most probably binding domains of the ORP, OLFD_CANFA (P30955).

Table 6

Gas detected	Frequency changes (Hz)			
	B1	B2	B3	Pb2

Trimethylamine (5.86 ppm)	5696	488	687	221
Dimethylamine(3.78 ppm)	3851	578	721	589
Ammonia (4.86 ppm)	1022	206	209	345
Acetone (7.21 ppm)	13	9	9	31
Formic acid (1.33 ppm)	161	56	85	97
Ethanol (4.68 ppm)	-5	6	-5	16
Formaldehyde (6.54 ppm)	-25	-22	-27	19

Peptide sequence of B1: DPDQRDC

Peptide sequence of B2: GDLESFC

Peptide sequence of B3: CFFLFFGD

Peptide sequence of Pb2: LFLSNLSFSDLCA

In Table 6, the numbers shown on each column under each peptide are frequency changes of the quartz crystal vibration frequency. Hence, the absolute value of the number is larger, and the sensitivity for the gas detected is larger. For the desirable detected gas, trimethylamine, all peptides B1, B2, and B3 show a much more sensitive response then the peptide Pb2 designed by other methods.

It will be apparent to those skilled in the art that various modifications and variations can be made to the structure of the present invention without departing from the scope or spirit of the invention. In view of the foregoing, it is intended that the present invention cover modifications and variations of this invention provided they fall within the scope of the following claims and their equivalents.